Content Dissemination by Pushing and Sharing in Mobile Cellular Networks: An Analytical Study

Xiaofei Wang*, Min Chen*, Zhu Han§, Ted "Taekyoung" Kwon* and Yanghee Choi*
*School of Computer Science and Engineering, Seoul National University, Seoul, Korea
§School of Electrical and Computer Engineering, University of Houston, Texas, USA

***Abstract*—The ever increasing traffic demand is a serious concern of mobile network operators, and the conventional pull-based (or request-based) communication model may not be able to handle this data explosion problem. To reduce the traffic load on cellular links for disseminating content, we propose to push the content to a subset of subscribers via cellular links, and to allow the subscribers to share the content via opportunistic local connectivity (i.e. Wi-Fi ad-hoc mode). We theoretically model and analyze how the content can be disseminated by both pushing via cellular links and sharing via Wi-Fi links, where handovers are modeled based on the multi-compartment model. We also formulate a mathematical framework to optimize the content dissemination, by which the trade-off between the dissemination delay and the energy cost is explored.**

I. Introduction

The data explosion problem in mobile cellular networks has become the most critical issue [1]. Mobile network operators (MNOs) seek to mitigate the traffic burden on their cellular links. As the link capacity enhancement in current mobile cellular networks (e.g., 3G and LTE) is unlikely to keep pace with the soaring traffic demand due to limited frequency spectrum, we should investigate this issue from other perspectives.

One of the outstanding trends in the Internet traffic is that increasingly more traffic is attributed to content-oriented applications and services. From this perspective, in addition to the traditional pull-based (request-based) communications, users (or applications) increasingly tend to subscribe to some pushing services from content providers (CPs), and the CPs push the content to subscribers as soon as the content is generated. For instance, the Really Simple Syndication (RSS) is one of the most popular pushing services, by which users can receive the newest photos, documents and video clips. Also `YouTube` provides some channel-based subscription service to push new and popular videos to users. Many applications in smart phones rely on push mechanisms as well. There are some studies to demonstrate the advantages of push-based models over pull-based models in various contexts (e.g., push-to-peer streaming [2] and mission-critical applications [3]).

From delay perspective, users may not always have to instantly access the content of interest as soon as the content is generated. Instead, some delay is tolerable depending on the users' daily lives and the content natures. For instance, a new music video is generated in the morning, but many people may watch it in the evening or even after some days. Also as reported in [4], when people download content files, there is a substantial disparity in the popularity of the files. That is, only a small portion of content files may be downloaded by a large number of users, which results in multiple users downloading the same content multiple times via cellular links redundantly [4] [5]. Therefore, it is attractive to exploit the affordable delivery delay in such a way that users can receive the content via non-cellular links (e.g., Wi-Fi). For instance, if a user who is to be pushed a file learns that another user who already got the file is nearby, they can "share" the file via Wi-Fi ad hoc connectivity.

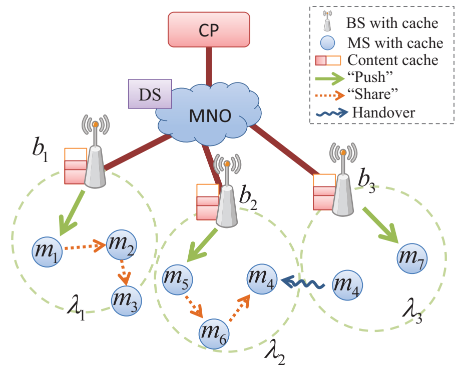


Fig. 1. An illustration of content dissemination by pushing via cellular links and by sharing via Wi-Fi links among MSs with handovers

From the above observations, we propose to use both "pushing" (over cellular links) and "sharing" over Wi-Fi links for the content dissemination to subscribers, which can reduce the traffic load on cellular links. The content can be of any type, such as news articles, stock information, advertisements, social events, weather forecasts, and video clips (which currently consumes more than a half of the whole mobile traffic [1]).

We simply illustrate how a file is disseminated by pushing and sharing in Fig. 1. Once a file (to which users have subscribed) is generated, the CP sends the file to a dissemination server (DS) in the MNO. The DS is in charge of disseminating the file to the subscribed users until its deadline (or the maximum tolerable delivery delay). The DS will deliver the file to the caching spaces of base stations (BSs), each of which then pushes the file to mobile stations (MSs) of the subscribed users via cellular links. Note that only a subset of MSs will receive the file by the pushing. If an MS with the file opportunistically gets in contact with another nearby MS without the file, they will set up a Wi-Fi connectivity to share the file. The opportunistic content delivery by these

978-1-4673-2433-5/12/$31.00 ©2012 IEEE

"social meets" has been extensively studied in the name of delay tolerant networks (DTNs) [6] [7] [8] [9]. We assume that every MS wakes up periodically with a low duty cycle to probe other MSs nearby for content "sharing" purposes.

The advantages of offloading the cellular traffic by the opportunistic user-to-user sharing have been discussed in prior studies [3] [9] [10] [11]. Let us compare pushing and sharing with other well-known strategies of content dissemination:

- **Pull-based Unicast** In the traditional pull-based delivery, the file of interest may be downloaded via cellular links as many times as the number of subscribers [4] [5]. Meanwhile, our proposed model leverages the social meets of users, to offload the redundant downloads from the cellular links to Wi-Fi links.
- **Broadcast/Multicast** When multiple users (in the same BS) wish to receive the same content, broadcasting (or multicasting) would be efficient. However, for broadcasting, the lowest bit rate is normally used to cover all the MSs in its cell, which reduces the efficiency substantially. Moreover, the reliability of the content delivery is still difficult to achieve. There are also security issues, such as overhearing in broadcast and complicated key management in multicast.

The focus of this paper is on how to coordinate the pushing and the sharing in the dissemination. Also, by using the multi-compartment model, we discuss how the content is disseminated among multiple cells with handovers. We further formulate an optimization framework for the dissemination performance, and explore the trade-off between the energy cost and dissemination delay. To the best of our knowledge, this is the first study to theoretically model and analyze the content dissemination across multiple cells in cellular networks based on pushing and sharing.

The rest of the paper is organized as follows. After reviewing the related work in Sec. II, we introduce the system model in Sec. III, and then detail the content dissemination process in a single cell and multiple cells in Secs. IV and V, respectively. We discuss how to optimize the system parameters in Sec. VI. Numerical results are shown in Sec. VII, followed by concluding remarks in Sec. VIII.

II. Related Work

The epidemic opportunistic content delivery in DTNs has been extensively studied these years. Zhang et al. [8] developed the differentiation-based model to study the delay of content delivery in DTNs. For the purpose of energy conservation, Li et al. [12] designed an opportunistic transmission scheme to control the content delivery in DTNs. However related studies on epidemic sharing suffer from slow start and long completion time, which motivates us to combine pushing and sharing to expedite the content dissemination.

How to offload traffic from cellular networks by opportunistic node-to-node sharing has become a hot research topic recently. Doppler et al. [13] exploited device-to-device communications as an underlay to LTE cellular networks for efficient content delivery. Also Whitebeck et al. [11] demonstrated the effectiveness of push-based offloading strategies based on practical mobility traces, but related analytic modeling was missing. The scalability and optimality of content offloading by exploiting user-to-user contacts were theoretically discussed in [9], where the balance between bandwidth resource and user mobility is studied as a social welfare maximization problem. Similarly, [10] utilized a number of mobile agents to disseminate contents, and solved the maximization of offloading utility as a knapsack problem. However the above studies were limited to single cell environments, while we consider multi-cell environments with handovers of MSs.

TABLE I
VARIABLES AND NOTATION OF THE SYSTEM MODEL

Variable	Explanation (default value in evaluation)
b_i	BS with id i
n	number of total BSs (20)
M_i	number of MSs in the area of b_i (1000)
m_k	a typical MS with index k
λ	average meeting rate of MSs in the cell (0.00001)
ϕ	energy consumption per delivery via Wi-Fi (1)
Φ	energy consumption per delivery via cellular link (4)
ρ	probing cost per time unit (0.001)
P_{init}	the amount of initial push (50)
P_{final}	the amount of final push (50)
$S(t)$	the function of number of updated MSs in the cell to time t
t^O	dissemination completion time with only pushing
t^*	dissemination completion time with both sharing and pushing
$C(t)$	accumulative cost function of MSs in the cell to time t
C^*	cost to disseminate content to all MSs in the cell
ℓ_{xy}	handover rate of MSs from BS b_x to BS b_y

III. System Model

We illustrate a dissemination scenario in Fig. 1, where there is one CP and one MNO with three BSs, b_1, b_2 and b_3. Each BS services multiple MSs in its cell who are interested in the CP's content. For example, m_1, m_2 and m_3 are within the cell of b_1. In this paper, we only focus on those MSs who have subscribed to the content from the CP, and thus ignore other ones. Even though a single CP may disseminate multiple files to the MSs periodically or concurrently, we focus a single file for sake of exposition. Also we do not consider MSs who may turn off during the dissemination. The notation and the default values are shown in Table 1.

As for "pushing", the CP first delivers a file for a particular group of MSs(its group identifier is needed) to the DS of the MNO. In a cellular network, the location management entity (LME) [14] keeps track of the locations of the MSs. Thus, along with the LME, the DS knows: (i) which MSs have subscribed the content, and (ii) which MSs of the group are serviced by each BS. Then the DS will dispatch the file to all the BSs that service the MSs. Each BS will initially push the content file to some of the MSs in its cell. For instance, BS b_1 will deliver the content to m_1 at the beginning. There can be different strategies regarding which MSs will be pushed first, but this is out of the scope of this paper (see [9] and [15] for details). In our paper we deploy a random strategy which we will describe later in Sec. IV-B.

As for "sharing", MSs will move with a certain mobility model. According to [6] [7] [8] and [9], the intervals between consecutive meets of any pair of MSs, called the inter-contact times (ICTs), are assumed to follow an exponential distribution. Also based on the measurements in [6], we assume that MSs at different places will have different mobility patterns and thus MSs at different BSs will have different mean rates of inter-contacts, denoted as λ_i for BS b_i, (also called meeting rate interchangeably). For instance, a park area will have a longer ICT than a subway station. Each MS periodically probes to check whether there is any nearby MS that holds the content being disseminated. We assume the MSs are synchronized and the probing is triggered with a sufficiently low duty cycle, say, during the first 5ms period in every second. (The energy consumption per time unit due to probing is denoted by ρ.) If there is, two MSs will share the content via ad-hoc Wi-Fi connectivity. For instance, m_1 occasionally meets m_2 and shares the content, and later m_2 meets and shares with m_3. If an MS obtained the content by either pushing or sharing, we say the MS is "updated".

Some MSs do not like to participate in carrying and sharing content with others due to security, privacy or cost issues. In this paper, we will exclude those MSs from the model. Note that related security and privacy issues in sharing can be handled by some prior work in opportunistic DTN such as [16] [17]. Also since the focus of this paper is to model and analyze how the content can be disseminated across multiple cells in a macro perspective, we assume that the content can be shared successfully via Wi-Fi during the meets with fairly high bit rates.

With the initial pushing and sharing, some MSs may not be able to obtain the content for a long time due to the limitation of the opportunistic sharing. Those MSs will request the final push from the BS, to be detailed later in Sec. IV-B.

IV. Content Dissemination in a Single Cell

In this section, we discuss the content dissemination within a single cell. For simplicity, we temporarily assume that for a certain amount of duration, the MSs will stay in a single cell and will not make handovers. We will discuss the case of multiple cells with handovers in Sec. V. As we consider a single BS in this section, we omit the BS's index i.

A. Content Dissemination by Sharing Only

We first focus on how the content is gradually disseminated to MSs over time t in a single cell by sharing only, where the number of MSs who are to receive the content is denoted by M. Let $S(t)$ be the state of the continuous-time Markov chain system, which indicates the number of MSs that have received the content until time t by sharing. We will obtain $S(t)$ from its derivative based on the similar methodology as used in [8] and [12]; thus we show only the main steps for the sake of simplicity.

Due to the synchronized probing among MSs, an MS, say m_k, will always be able to discover other MSs in the Wi-Fi range. During a short period, say Δt, the probability for m_k to get the content from any MS who already got the content within Δt, denoted by $\theta_{t,t+\Delta t}(m_k)$, can be calculated by

$$\theta_{t,t+\Delta t}(m_k) = 1 - \left(1 - \left(1 - e^{-\lambda\Delta t}\right)\right)^{S(t)}. \qquad (1)$$

Then summing this probability across all the MSs that have not received the content at time t, the current number of updated MSs after Δt, $S(t+\Delta t)$, can be calculated as

$$S(t+\Delta t) = S(t) + \sum_{k=1}^{M-S(t)} \theta_{t,t+\Delta t}(m_k), \qquad (2)$$

whose expectation is given by

$$E[S(t+\Delta t)] = E[S(t)] + (M - E[S(t)]) \cdot E[\theta_{t,t+\Delta t}(x)] \qquad (3)$$

We obtain the derivative of $E[S(t)]$ by letting $\Delta t \to 0$,

$$\begin{aligned} E[\overset{\bullet}{S}(t)] &= \lim_{\Delta t\to 0} \tfrac{E[S(t+\Delta t)]-E[S(t)]}{\Delta t} \\ &= \lim_{\Delta t\to 0} \tfrac{(M-E[S(t)])\cdot\left(1-\left(1-\left(1-e^{-\lambda\Delta t}\right)\right)^{S(t)}\right)}{\Delta t} \\ &= (M - E[S(t)])(\lambda \cdot E[S(t)]). \end{aligned} \qquad (4)$$

By solving the above ordinary differential equation (ODE), we finally obtain the function $S(t)$ by

$$S(t) = \frac{S(0)\,Me^{M\lambda t}}{M - S(0) + S(0)\,e^{M\lambda t}}. \qquad (5)$$

Note that if there is no MS who has the content at the beginning, i.e., $S(0) = 0$, $S(t)$ will always be zero. Therefore, the BS should push the file to at least one MS, i.e., $S(0) = 1$, and then the MS with the content will disseminate the file to other MSs by sharing. Thus, $S(t)$ starting with only a single seed will increase by sharing over time as

$$S(t) = \frac{Me^{M\lambda t}}{M - 1 + e^{M\lambda t}}. \qquad (6)$$

From Eq. (5) we can calculate the required delay, t_{req}, to disseminate the content to S_{des} MSs ($1 \le S_{des} \le M$) by,

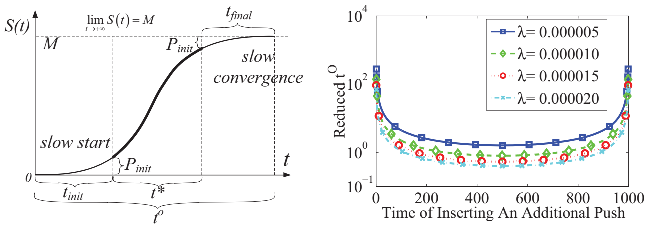
$$t_{req} = S^{-1}(S_{des}) = \frac{\log\left(\frac{S_{des}(M-S(0))}{S(0)(M-S_{des})}\right)}{M\lambda}. \qquad (7)$$

$S(t)$ from Eq. (5) in real domain cannot reach M in a finite time, which means $\lim_{t\to+\infty} S(t) = M$, and thus the dissemination completion time with only sharing, denoted by t^O, would be $t^O = S^{-1}(M) = +\infty$. However $S(t)$ actually takes integer values, so we define that the dissemination will be completed when $S(t) = M - \eta$, where η, $0 < \eta \ll 1$, takes a sufficiently small value, (e.g., $\eta = 1$). Then,

$$t^O = S^{-1}(M) \approx S^{-1}(M - \eta) = \frac{\log\left(\frac{(M-\eta)(M-1)}{\eta}\right)}{M\lambda}. \qquad (8)$$

B. Content Dissemination with Initial Push and Final Push

We illustrate $S(t)$ from Eq. (5) in Fig. 2(a), and we observe that the content dissemination only by sharing (starting with only a single seed) suffers from both a slow start and a slow convergence, due to the limitation of the opportunistic sharing. Therefore, we propose to increase the number of MSs who receive the content from the BS to reduce the delay..



(a) Content dissemination curve by sharing with a single seed (b) How much can one more push reduce delay? ($M = 1000$)

Fig. 2. How much can one more push accelerate the content dissemination?

In order to investigate "when" the BS should push the content for efficient dissemination, we evaluate how much dissemination completion time is reduced by pushing the content to one more MS at an arbitrary time (X-axis), as shown in Fig. 2(b). We observe that the additional pushing at the beginning and at the end can reduce the dissemination completion time most compared.

Therefore, we propose to disseminate content by three steps: (1) the BS pushes the file to a certain number of MSs at the beginning, denoted by P_{init}, which is actually $S(0)$ in Eq. (5); (2) the MSs will share the content file via opportunistic meeting; (3) when most of the MSs have received the content and there are still P_{final} not-yet-updated MSs, the BS finally pushes the file to them.

How to choose which MSs appropriately for initial pushing is out of the scope (see related work in [9] and [15]). Here we use a random strategy as follows: for each BS, the DS will calculate the optimal number of initial pushing P_{init} based on the environments in each cell (refer to the optimization framework in Sec. VI), and send the file to the BS, along with the ratio of $\frac{P_{init}}{M}$, the interest identifier and the dissemination deadline. Then each BS broadcasts a short message containing the information, and each MS who is interested in the content will reply to the BS with the probability of $\frac{P_{init}}{M}$ to confirm the initial pushing. In this way, the BS can push the file to P_{init} MSs probabilistically. At the deadline, the MSs who have not obtained the content will ask the BS to push the content to them finally. Each BS does not need to track the status of each MS and the dissemination progress.

Therefore, given the estimated P_{init} and P_{final}, the time to push the content to all the P_{final} MSs who have not received the content, denoted by t^*, is when the number of updated MSs $S(t)$ becomes $M - P_{final}$. Thus t^* becomes the practically dissemination completion time with both pushing and sharing.

Based on Eq. (7), we have

$$t^* = S^{-1}\left(M - P_{final}\right) = \frac{\log\left(\frac{(M-P_{final})(M-P_{init})}{P_{final}P_{init}}\right)}{M\lambda}. \quad (9)$$

Finally the content dissemination function $S(t)$ becomes a piece-wise function as follows,

$$S\left(t\right) = \begin{cases} P_{init} & t = 0, \\ \frac{P_{init} M e^{M\lambda t}}{M - P_{init} + P_{init} e^{M\lambda t}} & 0 < t < t^*, \\ M & t^* \le t. \end{cases} \quad (10)$$

C. Content Dissemination Energy Cost

The energy consumption is a critical issue for mobile networks because of the limited power supply of mobile devices. We mainly discuss the energy consumed at MSs for the content dissemination, which consists of:

- **Probing**: MSs periodically wake up with a sufficient duty cycle to detect whether there are nearby MSs with the content. We use ρ to denote the energy cost per time unit for probing, which is much smaller than those of receiving the content via a cellular link and sharing the content via a Wi-Fi link.
- **Pushing via cellular link**: We use Φ to denote the energy cost for receiving a file by BS's unicast via a cellular link.
- **Sharing via Wi-Fi link**: We use ϕ to denote the energy cost for transmitting and receiving a file from one MS to another via a Wi-Fi link. In practical, transmitting and receiving may consume different energy cost, but as they are will be just constants in our model, we hence assume the same value of them for simplicity, which will not affect our modeling. Thus the sharing of a file by Wi-Fi will cost 2ϕ. From the measurements in [18] and [19], Φ is greater than ϕ, and both are greater than ρ.

Therefore the accumulative energy cost for all MSs until time t can be then derived from Eq. (10) as follows:

$$C\left(t\right) = \begin{cases} \Phi P_{init} & t = 0, \\ \Phi P_{init} + 2\phi\left(S\left(t\right) - P_{init}\right) + M\rho t & 0 < t < t^*, \\ \Phi(P_{init}+P_{final})+2\phi(M-P_{init}-P_{final})+M\rho t^* & t^* \le t. \end{cases} \quad (11)$$

And after t^*, the energy cost for dissemination completion, denoted by C^*, can be calculated as

$$C^* = \Phi\left(P_{init} + P_{final}\right) + 2\phi\left(M - P_{init} - P_{final}\right) + M\rho t^*. \quad (12)$$

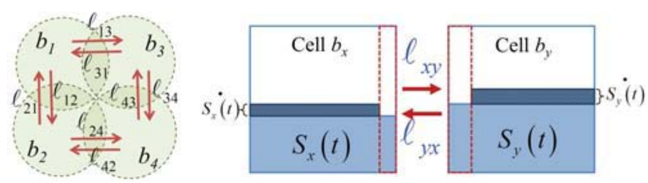
V. Content Dissemination for Multiple Cells

If we consider a number of BSs covering a large area, we should model the handovers among the cells, which strongly affect the content dissemination collectively. For instance, a BS covering a subway station will have many incoming and outgoing handover MSs, which either have the content or not. Thus we propose to adopt the multi-compartment model [20] to describe the content dissemination in a multi-cell scenario with handovers, based on the assumption that MSs' handovers follow a certain random process.

The multi-compartment model is commonly used in the biology fields (e.g., pharmacokinetics and biomedicine) to investigate the density of materials (e.g., drugs) in blood among different cells or parts of the organism, called compartments, and to track how the blood with the materials is circulating among compartments with some transition rates [20] [21]. These transitions from one compartment to another are similar to the handovers of MSs.

There have been some related studies for modeling handovers as a random process in [22] and [23]. According to these studies, the cell dwell time of an MS statistically follows a certain probability distribution (e.g., exponential distribution). We use the average rate of the random process model to represent the handover rate.

In an example scenario in Fig. 3(a), there are four cells b_1, b_2, b_3 and b_4, and between two neighbor BSs, the MSs are performing handovers in or out with a certain rate, denoted by ℓ_{xy}, which is defined as the probability that an MS moves from BS b_x to another BS b_y during a time unit. Note that handover rates can be obtained or estimated based on practical measurements by BSs and the LME in the MNO.



(a) A multi-cell scenario (b) How does the handover affect the content dissemination?

Fig. 3. Modeling the handovers in the content dissemination

The multi-compartment model is based not only on the handover rates but also on the number of the MSs, M_i, at each BS, b_i. We consider two kinds of scenarios for handovers to calculate M_i: (a) **non-steady-state scenario**, where the number of the MSs at each cell dynamically changes, for instance, a BS in a residential area during commuting time; (b) **steady-state scenario**, where the number of MSs at each cell can be assumed to be unchanged if the incoming handovers and outgoing handovers balance.

We also define the neighborhood set, Ω_i, which includes all neighbor BSs of b_i, for instance, $\Omega_1 = \{b_2, b_3\}$ in Fig. 3(a).

A. Non-steady-state Modeling of MSs in Multiple Cells

In non-steady-state scenarios, M_i of each BS b_i is dynamically changing; thus we use function $M_i(t)$ since M_i is changing over time t. Therefore, in a short period, its derivative, $\dot{M}_i(t)$, can be calculated based on the difference between incoming MSs and outgoing MSs as follows,

$$\dot{M}_i(t) = -M_i(t) \sum_{b_k \in \Omega_i} \ell_{ik} + \sum_{b_k \in \Omega_i} (\ell_{ki} M_k(t)). \tag{13}$$

Thus, for the n BSs, there will be n equations, which formulate a 1st-order linear homogenous ODE system. Referring to [24] and [25], the general solution is given by

$$M_i(t) = \sum_{z=1}^{n^2} A_z e^{B_z(t-C_z)}, \tag{14}$$

where the coefficients A_z, B_z and C_z are coefficient constants that can be calculated straightforward, but we will skip the details due to space limit (see related work in [24] and [25]).

B. Steady-State Modeling of MSs in Multiple Cells

When the BSs are in a steady-state, the incoming and outgoing MSs practically make no change to the number of MSs at each BS. Then $M_i(t)$ of any BS b_i will be fixed to a static number M_i, which simplifies Eq. (13) to,

$$-M_i \sum_{b_k \in \Omega_i} \ell_{ik} + \sum_{b_k \in \Omega_i} (\ell_{ki} M_k) = 0. \tag{15}$$

Therefore, n BSs will generate a linear system with n equations, which can be easily solved to get M_i of each BS in the steady-state scenario.

C. How Handovers Affect the Content Dissemination

From the previous two subsections, we obtain the number of MSs at each cell in either non-steady-state or steady-state scenario. Thus, along with the known handover rates, we analyze how the handovers affect the content dissemination among cells based on the multi-compartment model. Note that we change $S(t)$ to $\mathbb{S}(t)$ to describe the dissemination function with handovers in multi-cell scenarios.

As illustrated in Fig. 3(b), at an arbitrary time t, there are $\mathbb{S}_x(t)$ updated MSs in cell b_x and $\mathbb{S}_y(t)$ updated MSs in cell b_y, which are represented by the light blue solid rectangles. Then during a short period, there will be two types of MSs in the cell: (a) MSs who are performing handovers; (b) MSs who are sharing the content. Note that we assume that during the period, (a)-type MSs will not share the file, and (b)-type MSs will not perform handovers. Then the red dashed rectangles represent the (a)-type MSs, who move from one cell to another, and the dark blue shadowed rectangles represent the newly updated MSs during the period shared by (b)-type MSs.

In the non-steady-state scenario, considering those two types of MSs, the derivative function of $\mathbb{S}_i(t)$ of BS b_i can be extended based on Eq. (4) as follows,

$$\begin{aligned} &\dot{\mathbb{S}}_i(t) = \\ &(M_i(t) - \mathbb{S}_i(t)) \left(1 - \sum_{b_k \in \Omega_i} \ell_{ik}\right) \left(\lambda \left(1 - \sum_{b_k \in \Omega_i} \ell_{ik}\right) \mathbb{S}_i(t)\right) \\ &- \sum_{b_k \in \Omega_i} \ell_{ik} \mathbb{S}_i(t) + \sum_{b_k \in \Omega_i} (l_{ki} \mathbb{S}_k(t)). \end{aligned} \tag{16}$$

And for the steady-state scenario, the $M_i(t)$ becomes M_i.

Finally, there will be a complicated ODE system with n differential equations for modeling the content dissemination with both pushing and sharing in multi-cell scenario.

In the steady-state scenario, the number of MSs at each BS is constant; thus, the above ODE system is a 1st-order quadratic homogenous ODE system with constant coefficient, which is a Riccati type matrix differential equation system. Jodar et al. [26] discussed its closed analytical approximation solution. Also Darling [27] proposed to convert the Riccati matrix different equations to 2nd-order linear ODE system to obtain explicit solutions. In non-steady-state scenario, the ODE system becomes a 1st-order quadratic homogenous ODE system with variable coefficients, which is difficult to obtain its exact analytical solution, but can be approximated by the power series methodology (see [28]). Furthermore, the homotopy perturbation method can be also applied to obtain the approximation of $\mathbb{S}(t)$ (see [29]). Due to the limited space, we skip the details of the solving procedure.

Regarding the energy cost for content dissemination in multiple cells with handovers, based on the above $\mathbb{S}_i(t)$ in Eq.(16), we can also easily extend $C_i(t)$ in Eq. (11) and C_i^* in Eq. (12), which are denoted by $\mathbb{C}_i(t)$ and $\mathbb{C}_i^*$, respectively.

VI. Optimization Framework

From previous modeling of the content dissemination in a single cell and multiple cells with handovers, we discuss the optimization framework for the DS in the MNO to allocate the P_{init} and P_{final} to all BSs, in order to achieve the minimum dissemination completion time and energy cost.

A. Minimum Dissemination Completion Delay

From Fig. 2(a), the effective allocation of the number of initial pushing and final pushing becomes critical for accelerating the content dissemination procedure for a shorter completion time. Then the problem becomes that, at any BS, by given a specific upper bound of the number of MSs that are going to be pushed, P_{total}, how to find the optimal values of P_{init} and P_{final} to achieve the minimum dissemination completion time t^* referring to Eq. (9):

$$\begin{aligned} &\min_{P_{init},P_{final}} \{t^*\} \\ &\text{Subject to}: \quad P_{init} + P_{final} = P_{total}. \end{aligned} \tag{17}$$

We replace P_{final} by $P_{final} = P_{total} - P_{init}$, and find the minimum value by letting $\frac{\partial t^*}{\partial P_{init}} = 0$, so that the optimal value of P_{init} is found as $P_{init} = \frac{P_{total}}{2}$, which means that the BS should always equally allocate the number of initial pushing and that of final pushing so that the dissemination completion time t^* can be minimized, regarding a limited total number of pushing. Therefore, in the rest of the paper, we will just focus on the number of initial pushing, P_{init}, and consider $P_{final} = P_{init}$ by default. Note that the values of P_{init} and P_{final} should be less than the $\frac{M}{2}$.

B. Minimum Dissemination Completion Cost

Referring to the measurements in [18] and [19], Φ is several times larger than ϕ for one content delivery. If a BS pushes the content to more MSs via the cellular link in order to get a smaller t^*, it may consume more energy; otherwise if a BS pushes to less MSs inducing a larger t^*, it may also consume a large amount of probing energy over time. Thus we have the problem on how to find the optimal value of P_{init} to minimize the energy cost for completing the dissemination as follows,

$$\min_{P_{init}} \{C^*\}. \tag{18}$$

Based on Eq. (12), we use the similar method in the previous subsection to solve $\frac{\partial C^*}{\partial P_{init}} = 0$, and find the optimal P_{init} for the minimum C^* as,

$$P_{init} = \frac{M}{2} - \frac{\sqrt{M\lambda\left(M\lambda\left(\Phi - 2\phi\right) - 8\rho\right)\left(\Phi - 2\phi\right)}}{2\lambda\left(\Phi - 2\phi\right)}, \tag{19}$$

under the condition of,

$$\left(M\lambda\left(\Phi - 2\phi\right) - 8\rho\right)\left(\Phi - 2\phi\right) \geq 0. \tag{20}$$

Then we can obtain the minimum C^* referring to Eq. (12). When the condition in Eq. (20) equals to or less than 0, the optimal P_{init} with $P_{init} = \frac{M}{2}$ will lead to the minimum C^*.

In the multi-cell scenario, each BSs, b_i, can locally calculate the optimal P_{init_i} to minimize the energy cost $\mathbb{C}_i^*$, unless there is a limitation on the total number of the MSs being pushed among all BSs, P_{budget}, which is smaller than the sum of the local optimal values of P_{init_i}, that is

$$\sum_{\forall b_i} P_{init_i} < P_{budget} < \sum_{\forall b_i}\left(\arg\min_{P_{init_i}} \mathbb{C}_i^*\right). \tag{21}$$

With this constraint, the local optimization for each cell will not guarantee the minimum energy cost among all BSs. So the problem extends as

$$\begin{aligned} &\min_{\vec{P}_{init}} \left\{\sum_{\forall b_i} \mathbb{C}_i^*\right\} \\ &\text{subject to}: \sum_{\forall b_i} P_{init_i} < P_{budget}. \end{aligned} \tag{22}$$

It is hard to verify the convexity of $\mathbb{C}^*$ to P_{init}. So we will firstly approximate the above objective function based on the power series methodology (see [28]), and then carry out numerical analysis.

C. Conjunctive Minimization of Delay and Cost

Because CPs, MNOs and MSs all desire for both minimum delay t^* and cost C^*, we try to carry out overall optimization on both of them. Due to the different unit of time and energy, we bring a weight factor, w, to combine t^* and C^* conjunctively, which is also considered as the Pareto-optimality , and w indicates the emphasis on either the cost or delay. Thus, for a single cell, we have the following minimization problem:

$$\min_{P_{init}} \{t^* + w \cdot C^*\}. \tag{23}$$

We solve it by letting $\frac{\partial(t^* + w\cdot C^*)}{\partial P_{init}} = 0$, and then obtain the solution as:

$$P_{init} = \frac{M}{2} - \frac{\sqrt{\lambda\left(M^2\lambda\left(\Phi - 2\phi\right) - 8M\rho - 4w\right)\left(\Phi - 2\phi\right)}}{2\lambda\left(\Phi - 2\phi\right)}, \tag{24}$$

with a condition that

$$\left(M^2\lambda\left(\Phi - 2\phi\right) - 8M\rho - 4w\right)\left(\Phi - 2\phi\right) \geq 0. \tag{25}$$

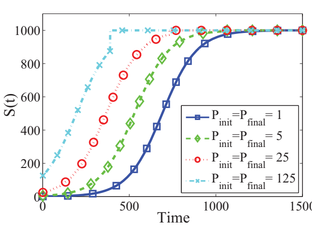
For the multi-cell scenario, if there is a constraint on the total amount of pushing, P_{budget}, the same to Eq. (21), each BS cannot push as it wants for local optimality between the delay and the cost; instead, all BSs must endeavor for the global optimization for following problem,

$$\min_{\vec{P_{init}}} \left\{ \sum_{b_i} (t_i^* + w \cdot \mathbb{C}_i^*) \right\} \qquad (26)$$
$$\text{subject to}: \sum_{\forall b_i} P_{init_i} < P_{budget},$$

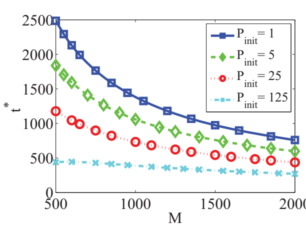
which is hard to find the minimum value by open-form solutions. Similar to previous subsection, we carry out approximation on the objective function for just numerical results.

VII. Evaluation

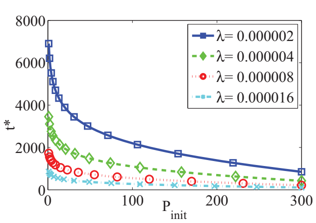
We simulate the continuous-time Markov system of our proposed model in Mathematica 8 along with the support of Matlab 2010 and Maple 14. For the purpose of evaluating the model realistically, we set the parameters with reasonable values based on previous mobility work in [7]: the meeting rate λ_i among MSs is from 0.0000001 to 0.0001 per second, and the number of MSs under one BS, M_i, is within the range from 300 to 3000. Also referring to [18] and [19], we set $\phi = 1$, $\Phi = 4$, and $\rho = 0.001$ per second by default. Note that the default values of the parameters are shown in Table 1.



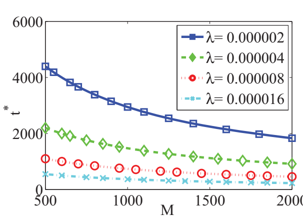
(a) $S(t)$ - different P_{init}



(b) t^*-different P_{init} with varying M



(c) t^*-different λ with varying P_{init}



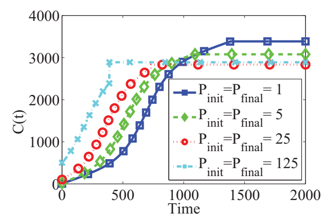
(d) t^*-different λ with varying M

Fig. 4. Evaluation on $S(t)$ and t^*

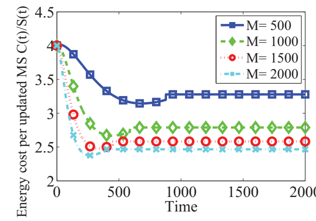
A. Content Dissemination within One Single Cell

The evaluation of the dissemination function $S(t)$ in Eq. (10) and the completion time t^* in Eq. (9) are shown in Fig. 4. From Fig. 4(a), when there is only one push ($P_{init} = 1$) at the beginning, the number of updated MSs starts to grow slowly, and converges to the dissemination completion slowly as well. When we increase the value of initial pushing, P_{init}, the dissemination procedure can be greatly shortened.

Regarding the completion time t^*, we observe that a cell with a small number of MSs will suffer from a large t^*, but a larger value for initial pushing P_{init} can reduce t^* dramatically as shown in Fig. 4(b). This indicates that when adjusting the values of initial pushing for the BSs, it is more beneficial to push more copies to small cells from the perspective of dissemination completion time. Fig. 4(c) and 4(d) both show that larger values of λ and M can significantly accelerate the dissemination and thus shorten t^*, because larger λ and M mean the higher probability that the MS can meet other MSs and thus be able to get the content by sharing. However, the benefit of increasing P_{init} is not significant when the meeting rate is high, as shown in Fig. 4(c).



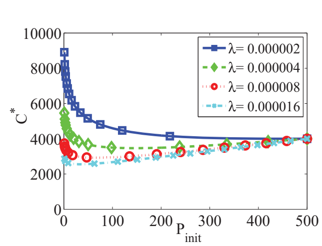
(a) Different P_{init}



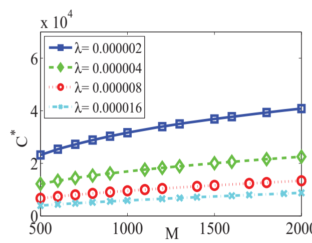
(b) Different M regarding $C(t)/S(t)$

Fig. 5. Evaluation of $C(t)$ - accumulative energy cost by time

The accumulative energy cost function of $C(t)$ in Eq. (11) is evaluated as shown in Fig. 5. From Fig. 5(a), we observe that the value of P_{init} has two-side impact on the $C(t)$: a small P_{init} ($P_{init} = 1$) will induce a long completion time, but the probing will consume a lot and thus $C(t)$ becomes quite large; however a large value of P_{init} ($P_{init} = 125$) can reduce t^* dramatically, but because of the more expensive energy cost for cellular links, it still consumes more C^* than that when $P_{init} = 25$. This falls into the optimization framework on C^* in Sec. VI-B, which we will discuss in later paragraphs. Furthermore, we calculate the energy cost per updated MS over time as shown in Fig. 5(b), and we discover that a large group will actually reduce the energy cost for each individual MS due to sharing.



(a) Different λ with varying P_{init}

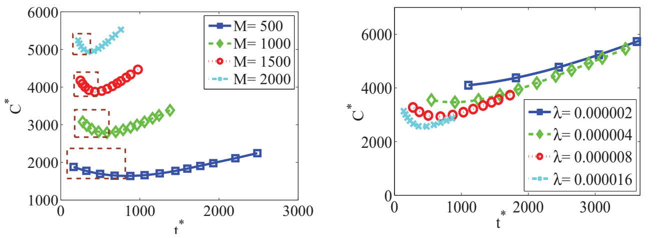


(b) Different λ with varying M

Fig. 6. Evaluation of C^* - the energy cost for dissemination completion

The evaluation on the energy cost for dissemination completion C^* in Eq. (6) is shown in Fig. 6. The relationship between P_{init} and C^* in Fig. 6(a) reflects our optimization framework in Sec. VI-B; P_{init} can be adjusted for a minimized C^* under the condition in Eq. (20). In the case that the condition is not satisfied ($\lambda = 0.000002$ in Fig. 6(a)), the optimal P_{init} for minimizing C^* will be $\frac{M}{2}$. Also when λ is larger, the optimal

P_{init} for minimum C^* is smaller. And also from Fig. 6(b), a higher meeting rate λ means more frequent social sharing via Wi-Fi, and it can significantly reduce the C^*, due to the lower energy cost of Wi-Fi links.

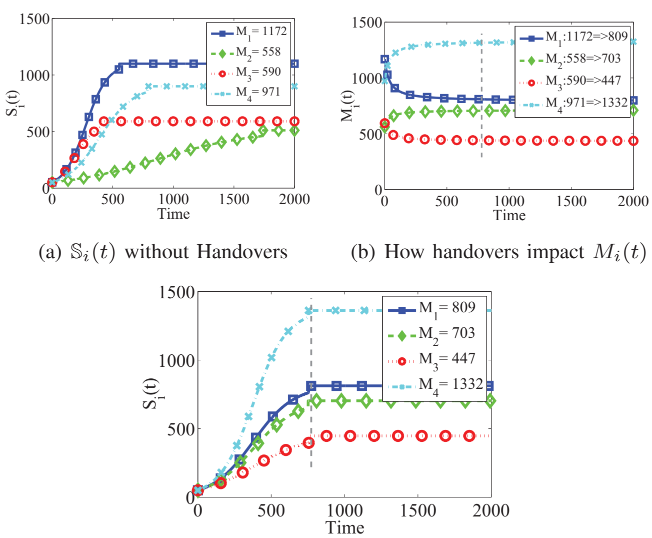


(a) Balance between C^* and t^* with different M (b) Balance between C^* and t^* with different λ

Fig. 7. Trade-off between C^* and t^* for completing the dissemination

The trade-off between C^* and t^* is explored in Fig. 7(b) and 7(a), when we adjusting P_{init} with different numbers of MSs and meeting rates. there is always a valley in the C^*-t^* curve, where C^* gets minimized (referring to Eq. (18)). The part of the curve on the left of the valley, where C^* and t^* are in an inverse relation, defines the boundary of the achievable delay-energy region (emphasized within the dashed rectangles) when P_{init} is higher than $\arg\min_{P_{init}}\{C^*\}$. This reflects the Pareto-optimal between C^* and t^* discussed in Sec. VI-C, and depending on the weight factor w, it is easy to find an optimal balance between C^* and t^* within the rectangle areas. On the right part of the curve, when P_{init} is not sufficiently large, the system will suffer from both high energy cost and long dissemination completion time.

B. Content Dissemination within Multiple Cells



(a) $\mathbb{S}_i(t)$ without Handovers (b) How handovers impact $M_i(t)$

(c) $\mathbb{S}_i(t)$ with handovers

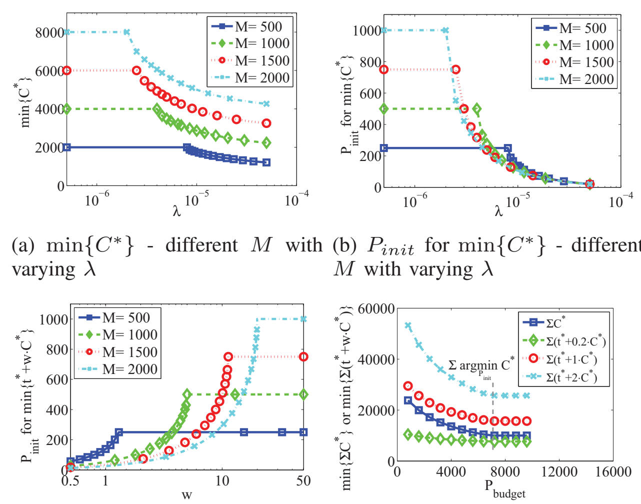
Fig. 8. Content Dissemination in Multi-cell Scenario with Handovers

For investigating the content dissemination in multi-cell scenario with handovers, we evaluate the MNO network in Fig. 3(a) as a typical example. At the beginning, for M_i and λ_i for each cell, we randomly assign practical values as introduced previously. Also the handover rates are set between 0.01 to 0.2 randomly, because the handover rates are not too high in real measurements [22] and [23].

We firstly plot $\mathbb{S}_i(t)$ of each cell without applying the handover rates as shown in Fig. 8(a). We can see that each BS completes the content dissemination separately, regardless of either the very slow dissemination of b_2 in green color (the diamond dashed curve) with $M_2 = 558$, or the very fast one of b_3 in red color (the circle dotted curve) with $M_3 = 590$.

Then we apply the handover rates to the model and examine $M_i(t)$ as shown in Fig. 8(b). Each BS changes the number of MSs due to the handovers of the MSs, and finally $M_i(t)$ converges to a steady-state around 520 seconds. Note that we approximately assume the steady-state when the change of $M_i(t)$ per second is small than 1. The corresponding plot of $\mathbb{S}_i(t)$ is shown in Fig. 8(c), and we can see the BSs complete the dissemination at the same time around 783 seconds. This is mainly because when MSs are performing handovers, some of them carry the content but the other do not; each cell will then exchange its both not-yet-updated MSs and updated MSs with its neighbor cells. The cells, which originally disseminate content fast, will "help" those who suffer from slow dissemination. Therefore, $\mathbb{S}_i(t)$ of BSs together grow and finally complete with same t^* in a harmonized manner.

C. Optimization Framework



(a) $\min\{C^*\}$ - different M with varying λ (b) P_{init} for $\min\{C^*\}$ - different M with varying λ

(c) P_{init} for $\min\{t^* + w \cdot C^*\}$ - different M with varying w (d) $\min\{\Sigma\mathbb{C}^*\}$ and $\min\{\Sigma(t^* + w \cdot \mathbb{C}^*)\}$ - different w with varying P_{budget}

Fig. 9. Optimization of C^*, $t^* + w \cdot C^*$, and $\Sigma(t^* + w \cdot \mathbb{C}^*)$

The minimization of C^* in a single cell is shown in Fig. 9(a) , 9(b) and 9(c). Note that the X-axis is in log scale. We see that with a larger meeting rate λ, the BS can adjust P_{init} to a smaller value for getting the minimum C^*. But when λ goes smaller below a boundary (referring to the condition in Eq. (20)), the BS will only have to set P_{init} to $\frac{M}{2}$ for the minimum C^*, which means to push the content to all of its MSs. The Pareto-optimality between the delay and energy cost is evaluated in Fig. 9(c), which indicates if the MNO system emphasizes more on the energy cost (a higher value of w), P_{init} should be set to a higher value until $\frac{M}{2}$. Fig. 9(d) shows the evaluation on the P_{budget}-constrained optimization in the multi-cell scenario (20 cells with reasonable parameter-s). Depending on the boundary condition in Eq. (21), when P_{budget} is sufficient, BSs can freely adjust P_{init_i}'s values individually for both local and global minimum cost; when P_{budget} is not enough, the minimum energy cost increases. Also $\Sigma(t_i^* + w \cdot \mathbb{C}_i{}^*)$ follows the same trend. Note that when P_{budget} is quite small, BSs will have small P_{init_i}, so BSs will mostly rely on the ICT-based sharing, and thus suffer from high energy cost and large delay.

VIII. Conclusion

In this paper, we proposed to reduce the traffic load on cellular links by coordinating pushing and sharing for disseminating delay-tolerant content. Content dissemination can be adaptively accelerated or decelerated to satisfy performance requirements by adjusting the initial and final pushing rates. The multi-compartment model can be adopted for modeling the content dissemination among multiple cells with handovers in cellular networks. The proposed optimization framework can be used by MNOs to control the pushing strategy for the objectives such as the minimum delay or minimum cost.

The lessons from the analytic studies are summarized as follows: pushing more copies to cells with the fewer MSs can be more beneficial for reducing the completion delay (Fig. 4(b)); the more users participate in sharing, the more energy saving can be achieved due to the sharing (Fig. 5(b)); the completion delay and energy cost exhibit an inverse relation, which reflects the Pareto-optimality when the required completion delay is small (the dashed boxes in Fig. 7(a)); if the requirement of completion delay is long, the energy cost of neighborhood monitoring will be overwhelming as shown in Figs. 7(a) and 7(b); the handovers among cells mix the MSs with or without the content, which implies a balance of overall completion delays among cells, and hence BSs can finish the content dissemination to their MSs with almost similar delays (Fig. 8(c)). In the future, we will extend the model for more practical scenarios such as heterogeneous mobility and transmission failure probability.

Acknowledgment

This research was supported by the KCC(Korea Communications Commission), Korea,under the R&D program supervised by the KCA(Korea Communications Agency) (KCA-2012-11-911-05-002)

References

[1] CISCO, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update , 2010-2015," CISCO, Tech. Rep., 2011.

[2] K. Suh, C. Diot, J. Kurose, L. Massoulié, C. Neumann, D. Towsley and M. Varvello, "Push-to-Peer Video-on-Demand system: design and evaluation," *IEEE JSAC*, no.9, pp.1706–1716, 2007.

[3] F. Chen, M. Johnson, A. Bar-Noy, I. Fermin and T. La Porta, " Proactive Data Dissemination to Mission Sites," *SECON*, 2009.

[4] Y. Chen, L. Qiu, W. Chen, L. Nguyen and R. Katz, "Clustering Web Content for Efficient Replication," *ICNP*, 2002.

[5] M. Cha, H. Kwak, P. Rodriguez, Y. Y. Ahn and S. Moon, "I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System," *ACM Internet Measurement Conference*, 2007.

[6] T. Karagiannis, J. Y. Le Boudec and M. Vojnovic, "Power Law and Exponential Decay of Inter Contact Times between Mobile Devices," in Proceedings of *ACM MobiCom*, 2007.

[7] R. Groenevelt, "The Message Delay in Mobile Ad-hoc Networks," *Performance Evaluation*, no. 1-4, pp. 210–228, 2005.

[8] X. Zhang, G. Neglia, J. Kurose and D. Towsley, "Performance Modeling of Epidemic Routing," *Computer Networks*, no. 10, pp. 2867–2891, 2005.

[9] S. Ioannidis, A. Chaintreau and L. Massoulie, "Optimal and Scalable Distribution of Content Updates over A Mobile Social Network," in Proceedings of *INFOCOM*, pp. 1422–1430, 2009.

[10] Y. Li, G. Su, P. Hui, D. Jin, L. Su and L. Zeng, "Multiple Mobile Data Offloading Through Delay Tolerant Networks," in *ACM Mobicom, CHANTS Workshop*, 2011.

[11] J. Whitbeck, M. D. Amorim, Y. Lopez, J. Leguay, V. Conan, "Relieving the Wireless Infrastructure: When Opportunistic Networks Meet Guaranteed Delays," in *IEEE WOWMOM*, 2011.

[12] Y. Li, Y. Jiang, D. Jin, L. Su, L. Zeng and D. Wu, "Energy-efficient Optimal Opportunistic Forwarding for Delay-Tolerant Networks," *IEEE Transactions on Vehicular Technology*, no. 9, pp. 4500–4512, 2010.

[13] K. Doppler, M. Rinne, C. Wijting, C. Ribeiro and K. Hugl, "Device-to-device communication as an underlay to LTE-advanced networks," *IEEE Communications Magazine*, no. 12, pp. 42–49, Dec, 2009.

[14] 3GPP, 3GPP TS 23.012, Location Management Procedures, 2011.

[15] B. Han, P. Hui, V. S. A. Kumar, M. V. Marathe, J. Shao and A. Srinivasan, "Mobile Data Offloading through Opportunistic Communications and Social Participation," *IEEE Transactions on Mobile Computing*, 2011.

[16] A. Shikfa, M. Onen and R. Molva, "Privacy in Content-Based Opportunistic Networks," *International Conference on Advanced Information Networking and Applications Workshops*,2009.

[17] A. Kate, G. M. Zaverucha, U. Hengartner, "Anonymity and Security in Delay Tolerant Networks," in *SecureComm*, 2007.

[18] N. Balasubramanian, A. Balasubramanian and A. Venkataramani, "Energy Consumption in Mobile Phones: A Measurement Study and Implications for Network Applications," in Proceedings of *ACM IMC*, 2009.

[19] S. Bohacek, "3G v.s WIFI Radio Energy with YouTube downloads," Tech. Report, 2009.

[20] A. Locker, "Multi-Compartment Models and Their Analysis," in *International Journal Of Clinical Pharmacology Therapy And Toxicology*, 1972.

[21] A. Iggidr, "Multi-compartment Models," *Discrete and Continuous Dyanmical Systems Supplement*, pp. 506–519, 2007.

[22] M. Zonoozi and P. Dassanayake, "User Mobility Modeling and Characterization of Mobility Patterns," *IEEE JSAC*, vol.15, pp.1239-1252, 1997.

[23] N, Hegde and K. Sohraby, "On the Impact of Soft Hand-off in Cellular Systems," in Proceedings of *ACM SIGMETRICS*, pp.178–187, 2000.

[24] D. Keffer, "An Analytical Method for Solving Systems of Linear nth-order ODEs," Lecture, 1999.

[25] S. Khorasani and A. Adibi, "Analytical Solution of Linear Ordinary Differential Equations by Differential Transfer Matrix Method," *Electronic Journal of Differential Equations*, vol.79, pp.1–18, 2003.

[26] L. Jodar and E. Navarro, "Closed Analytical Solution of Riccati Type Matrix Differential Equations," *Indian Journal of Pure Application Mathematics*, vol.23, no.3, pp.185–187, 1992.

[27] R. W. R. Darling, "Converting Matrix Riccati Equations To Second-Order Linear Ode," *SIAM Rev.*, vol.39, no.3, pp.508–510, 1997.

[28] N. Guzel and M. Bayram, "Power Series Solution of Non-linear First Order Differential Equation Systems," *Trakaya University Journal of Science*, vol.6, no.1, pp.107-111, 2005.

[29] H. Aminikhah and J. Biazar, "A New Analytical Method for System of ODEs," *Numerical Methods for Partial Differential Equations*, vol.26, no.5, pp.1115–1124, 2010.